



Semantic Web Technologies I

Lehrveranstaltung im WS11/12

Dr. Elena Simperl

PD Dr. Sebastian Rudolph

M. Sc. Anees ul Mehdi

AIFB Linked Data

•Dr. Elena Simperl

XML und URIs

Einführung in RDF

RDF Schema

Logik - Grundlagen

Semantik von RDF(S)

SPARQL - Syntax und Intuition

Semantik von SPARQL

OWL - Syntax und Intuition I

OWL – Syntax und Intuition II

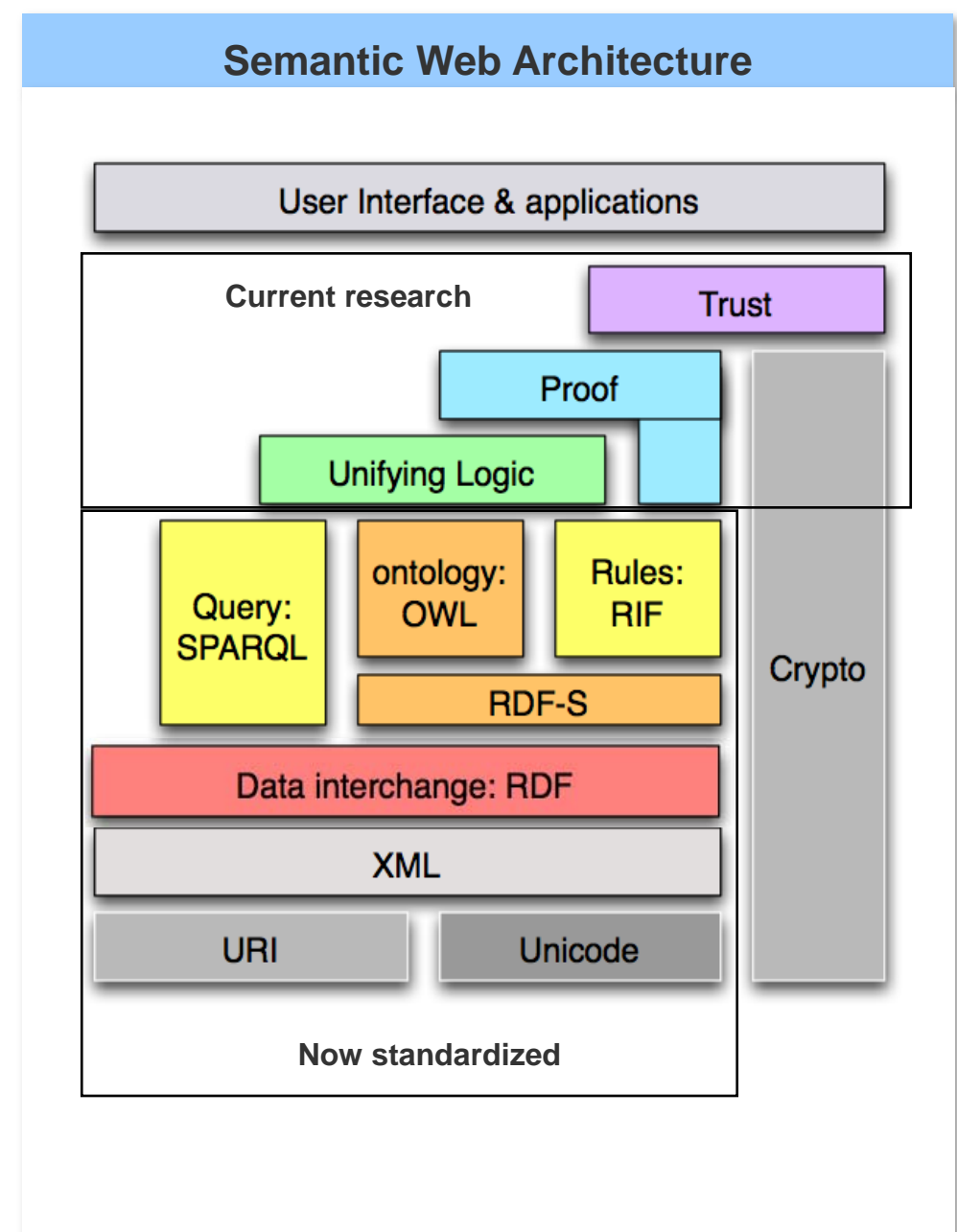
OWL - Semantik und Reasoning

Ontology Engineering

Linked Data

Konjunktive Anfragen / Einführung Regelsprachen

Anwendungen



<http://sild.cs.vu.nl/>

**SLIDES DUE TO HARTH, HOGAN,
KOTOULAS AND URBANI – SCALABLE
INTEGRATION OF LINKED DATA @
ISWC2011**

INTRODUCTION

Motivation



- With increased use of computers more and more data is being stored
 - Organisations rely on data for business decisions
 - Data drives policy decisions in government
 - Individuals rely on data from the Web for information and communication
- Data volumes explode
 - More and more data available on the Web is represented in Semantic Web standards
 - Linking Open Data (LOD) initiative
- Semantic Web technologies facilitate the integration of data from multiple sources
- Combining data from multiple sources enables insights

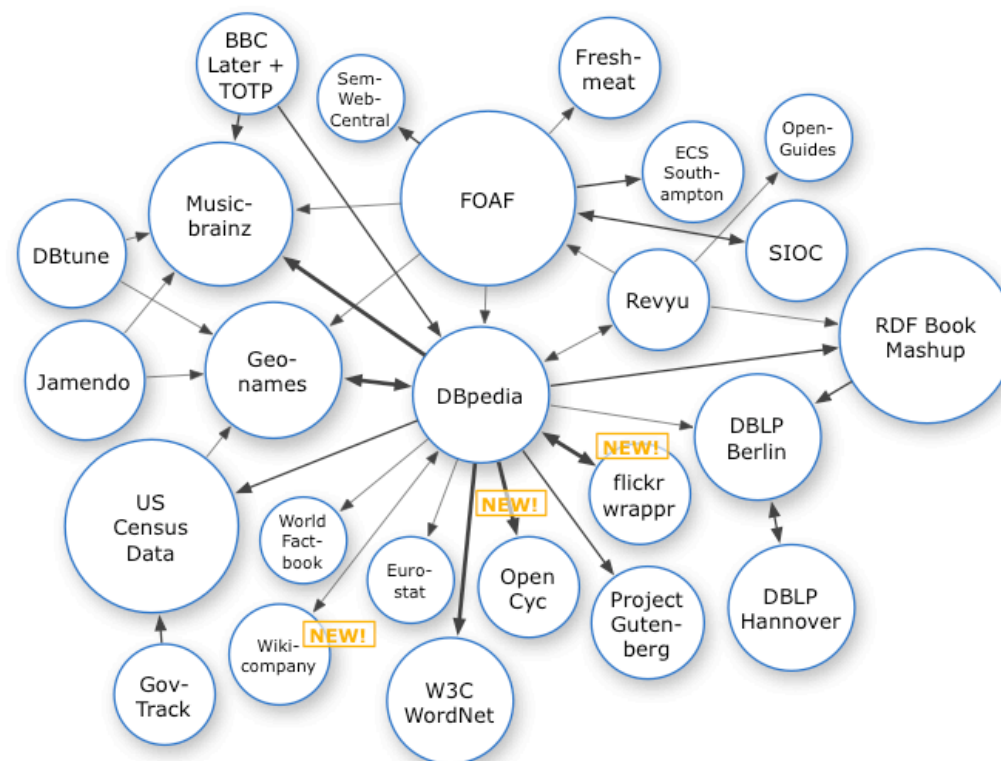
Linked Data Now!



http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html

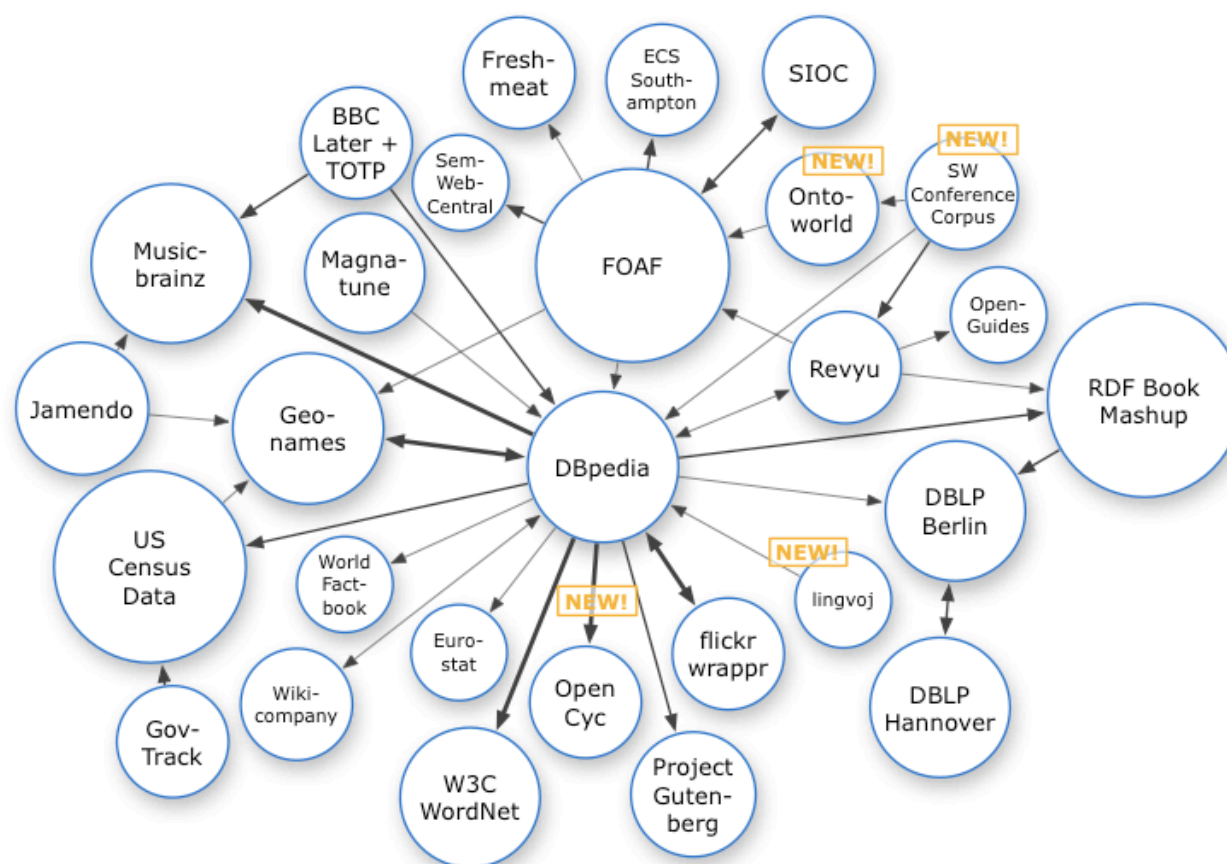
Linked Data on the Web

2007-10



Linked Data on the Web

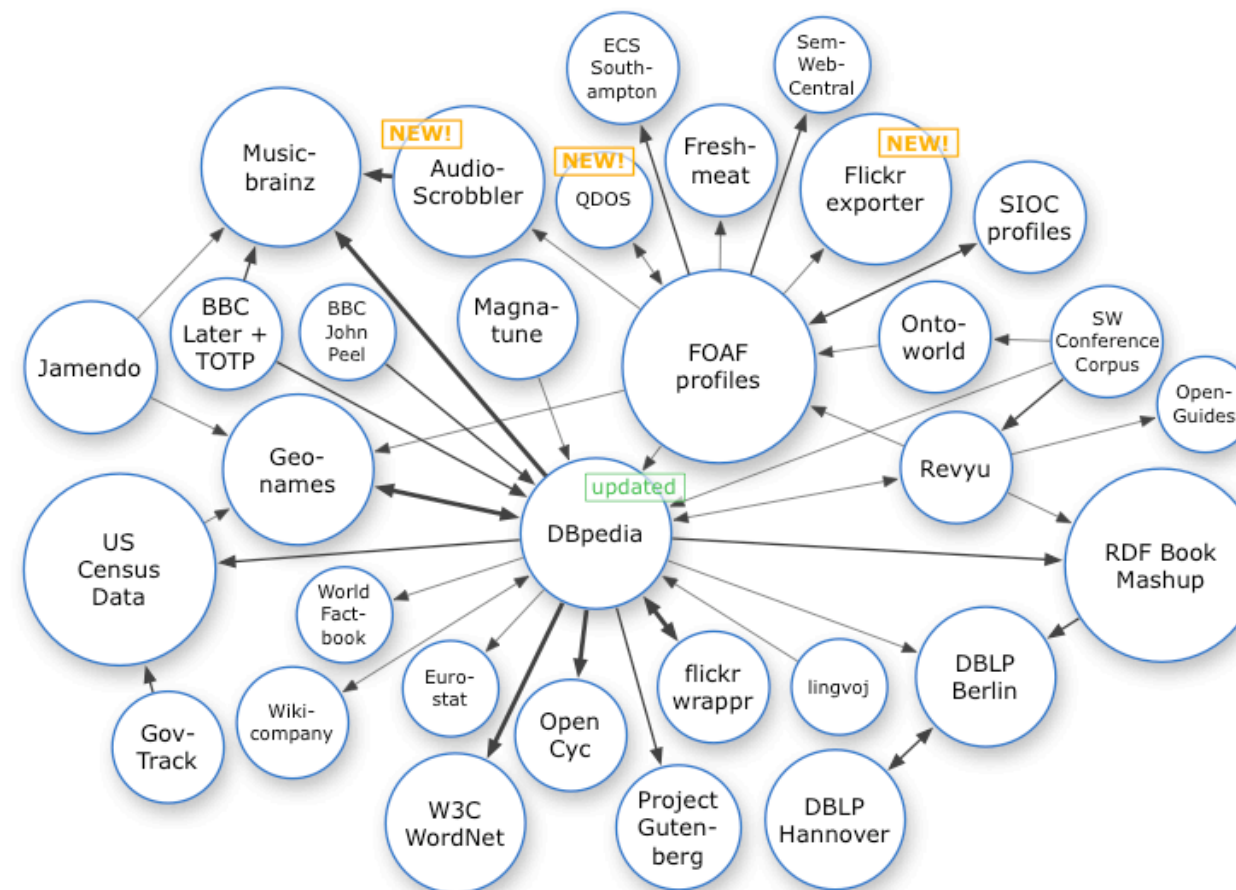
2007-11



Linked Data on the Web



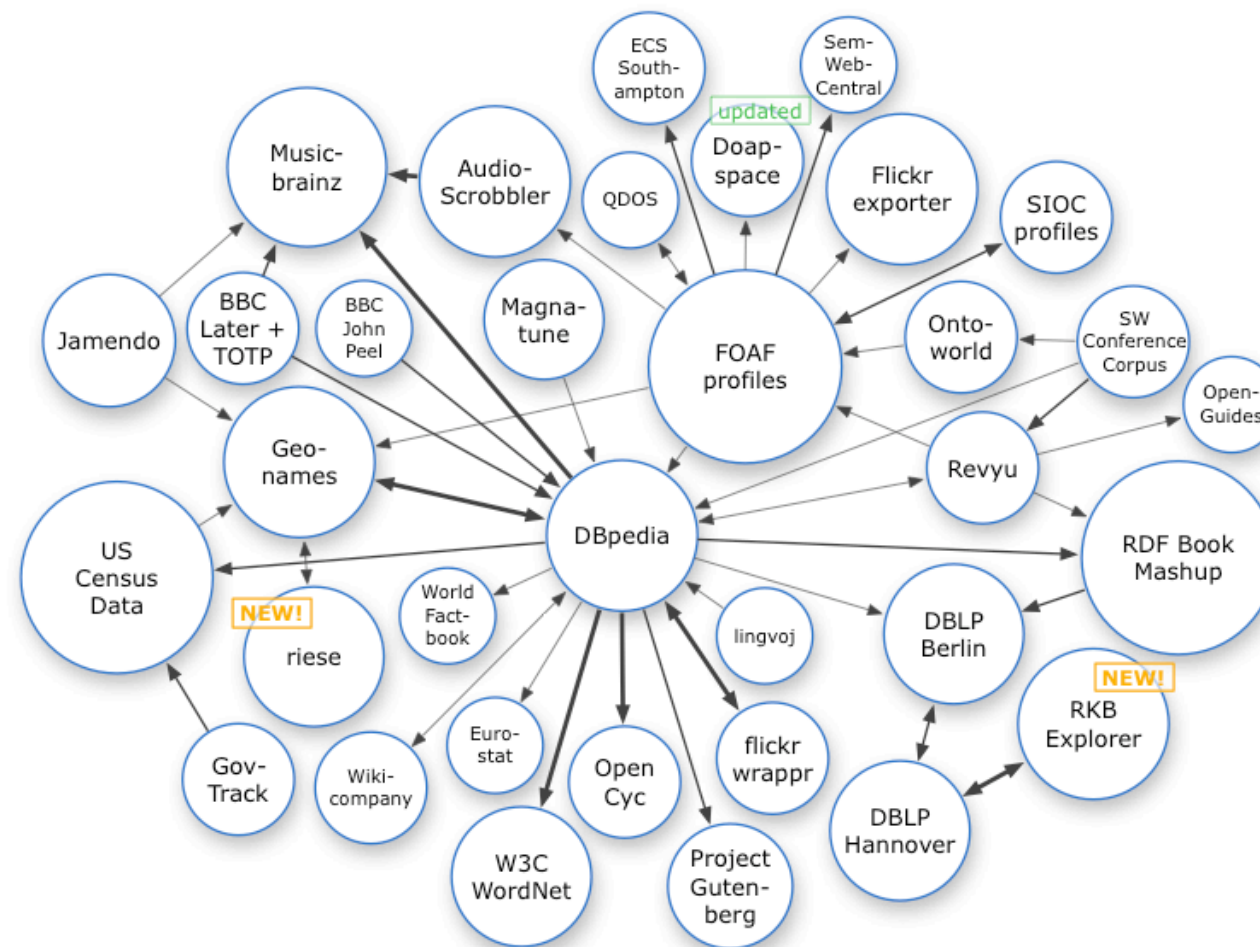
2008-02



Linked Data on the Web



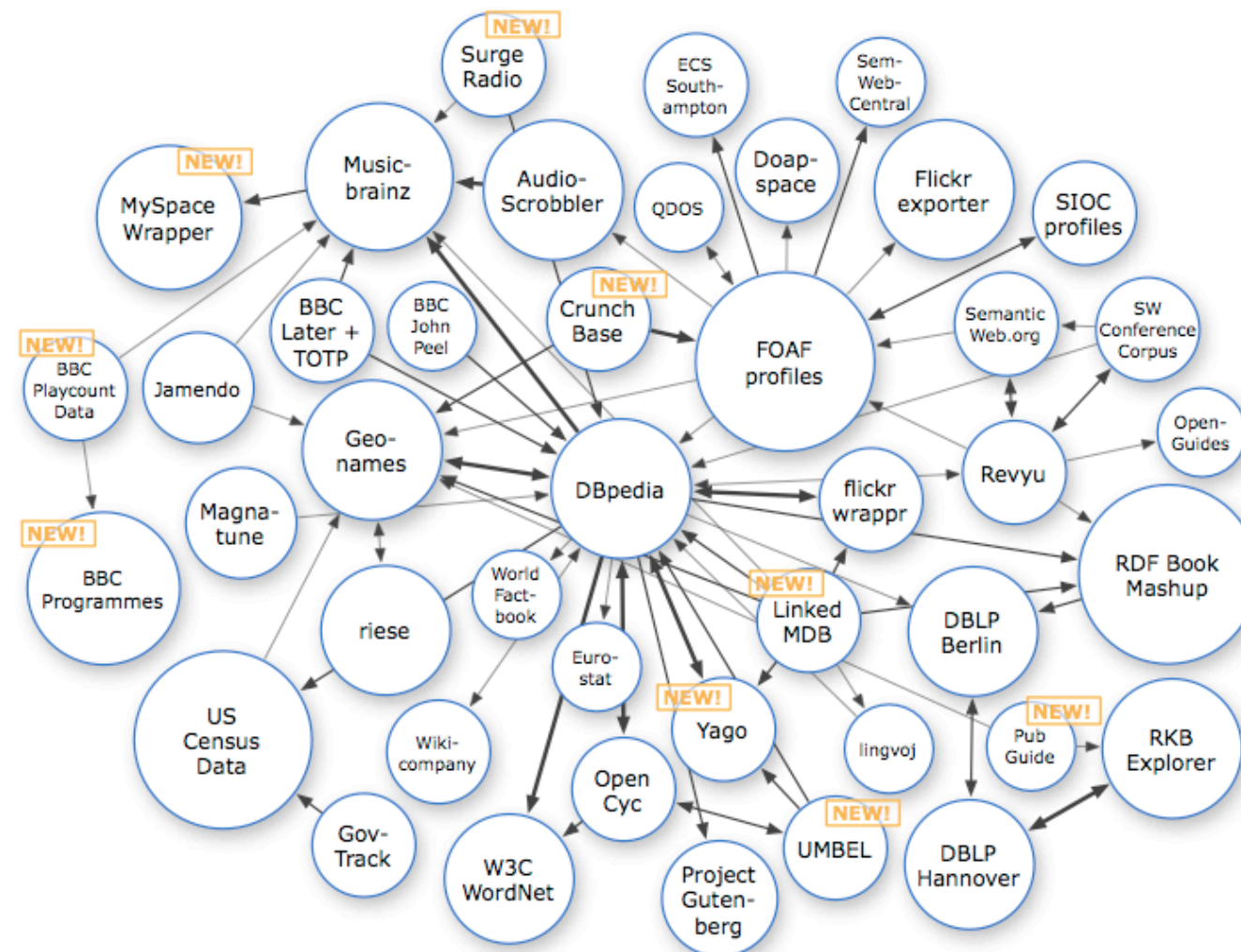
2008-03



Linked Data on the Web



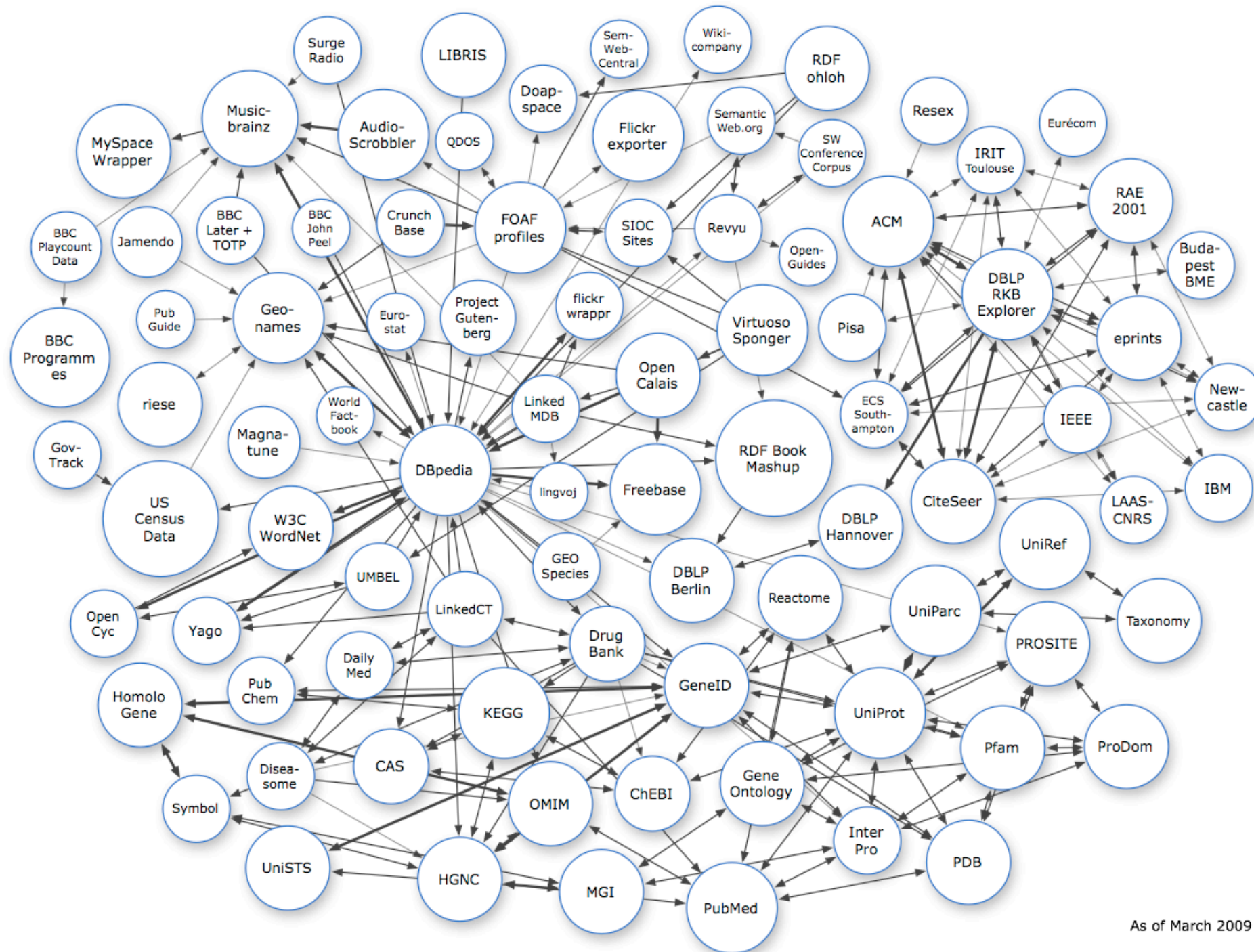
2008-09



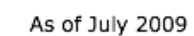
As of September 2008

Linked Data on the Web

2009-03

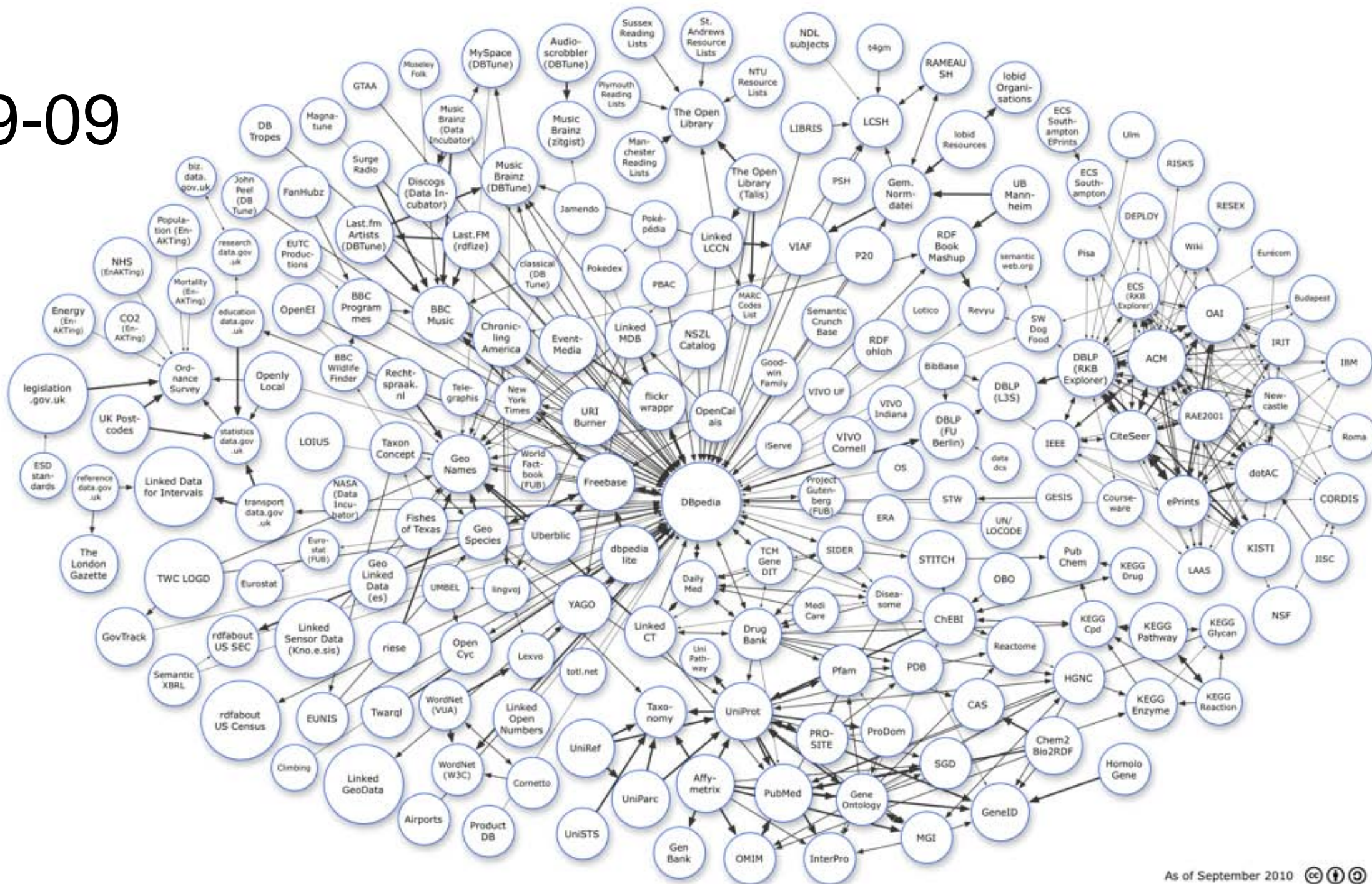


As of March 2009



Linked Data on the Web

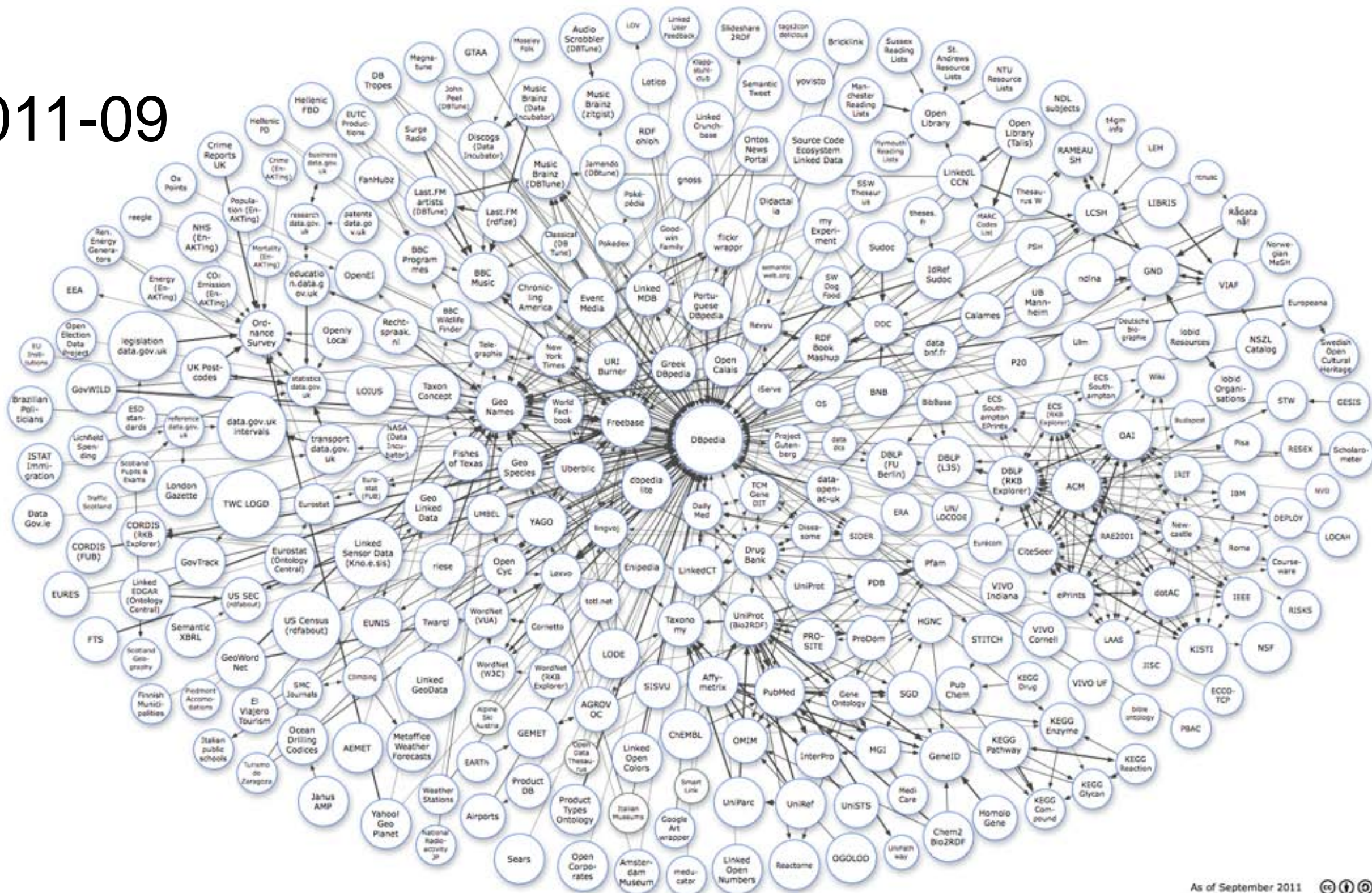
2009-09



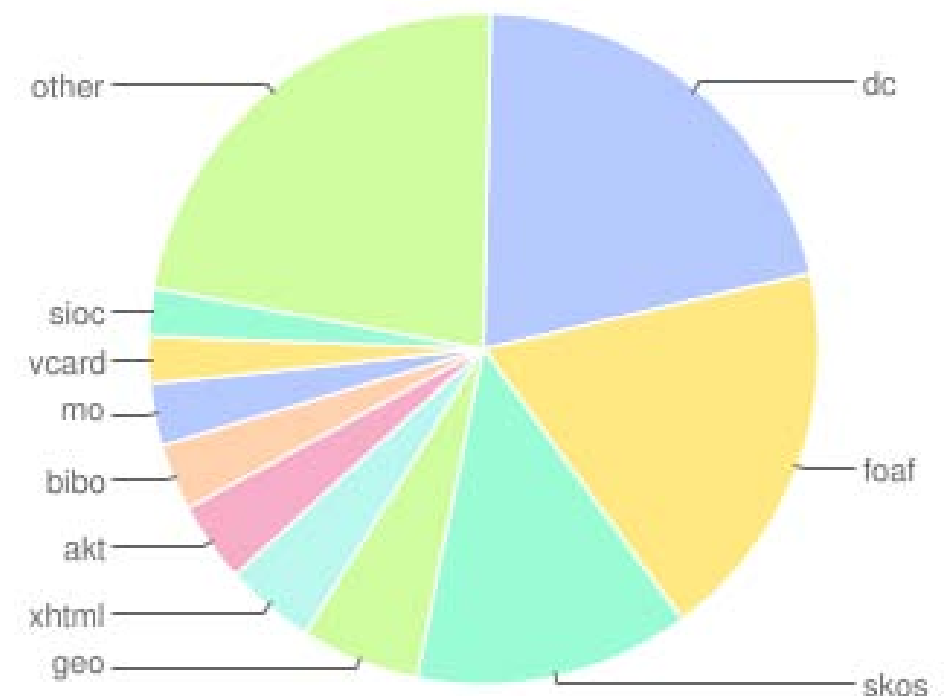
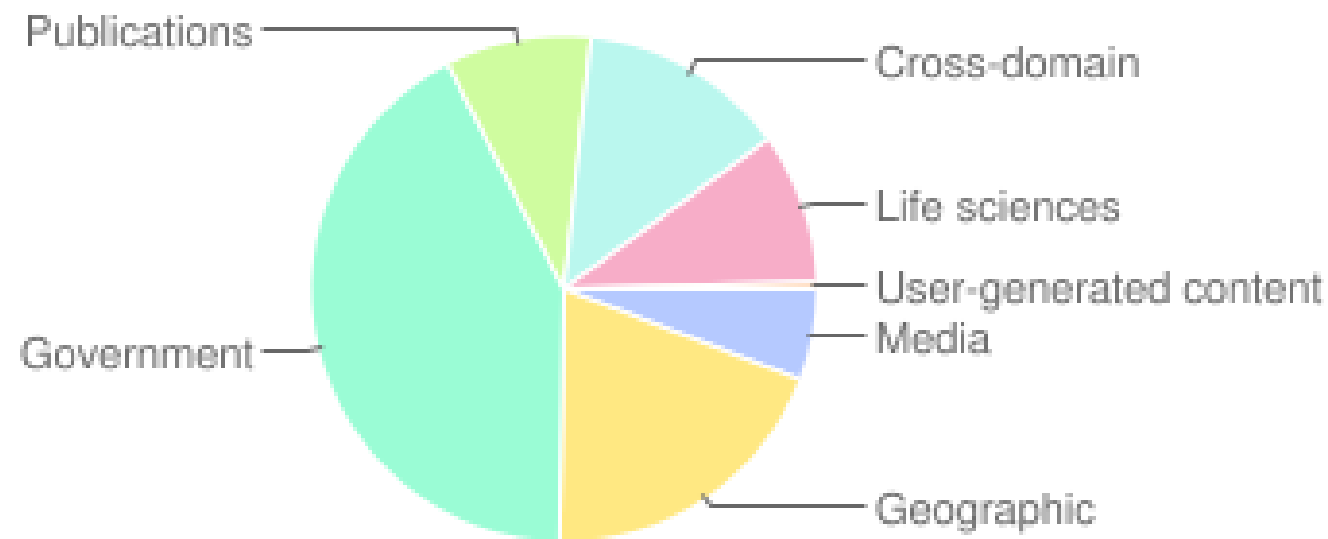
Linked Data on the Web

AIFR 

2011-09



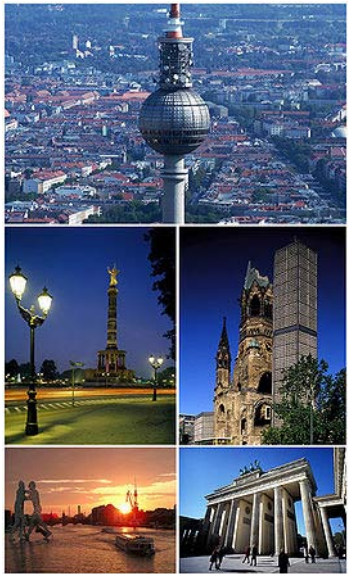
Data and ontologies in the Linked Open Data Cloud



<http://www4.wiwiiss.fu-berlin.de/lodcloud/state/>
(September 2011)

Scenario overview

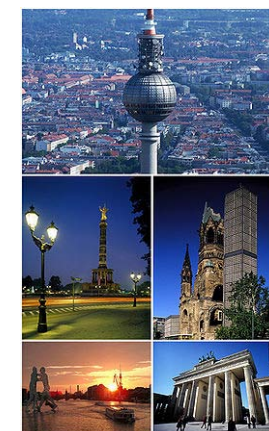
- Semantic technologies facilitate access to data
 - Q: data about Berlin?
 - Q: famous people that died in Berlin?
 - Q: data about Hegel?
 - Q: Hegel's publications?
 - Q: data about Marlene Dietrich?
 - Q: Dietrich's songs?



DBpedia



- Linked Data version of Wikipedia
- Scripts that extract data (text, links, infoboxes) from Wikipedia
- Published as Linked Data
- Interlinking hub in the Linked Data web
- Berlin
 - <http://dbpedia.org/resource/Berlin>
- Hegel
 - http://dbpedia.org/resource/Georg_Wilhelm_Friedrich_Hegel
- Marlene Dietrich
 - http://dbpedia.org/resource/Marlene_Dietrich



BBC Music

AIFB 

- Data about BBC (radio) programmes, artists, songs...
- Combination of BBC-internal data (playlists), MusicBrainz (artists, albums), Wikipedia (artists)
- Underpinning the BBC Music website
- Data published according to Linked Data principles
- Marlene Dietrich
 - <http://www.bbc.co.uk/music/artists/191cba6a-b83f-49ca-883c-02b20c7a9dd5>



Virtual International Authority File (VIAF)

AIFB 

- Joint project of national libraries and related organisations
 - 21 institutions, among them the Deutsche Nationalbibliothek
- Provide access to “authority files”
- Matching and interlinking collections from participating institutions
- Hegel
 - <http://viaf.org/viaf/89774942>
- Marlene Dietrich
 - <http://viaf.org/viaf/97773925>



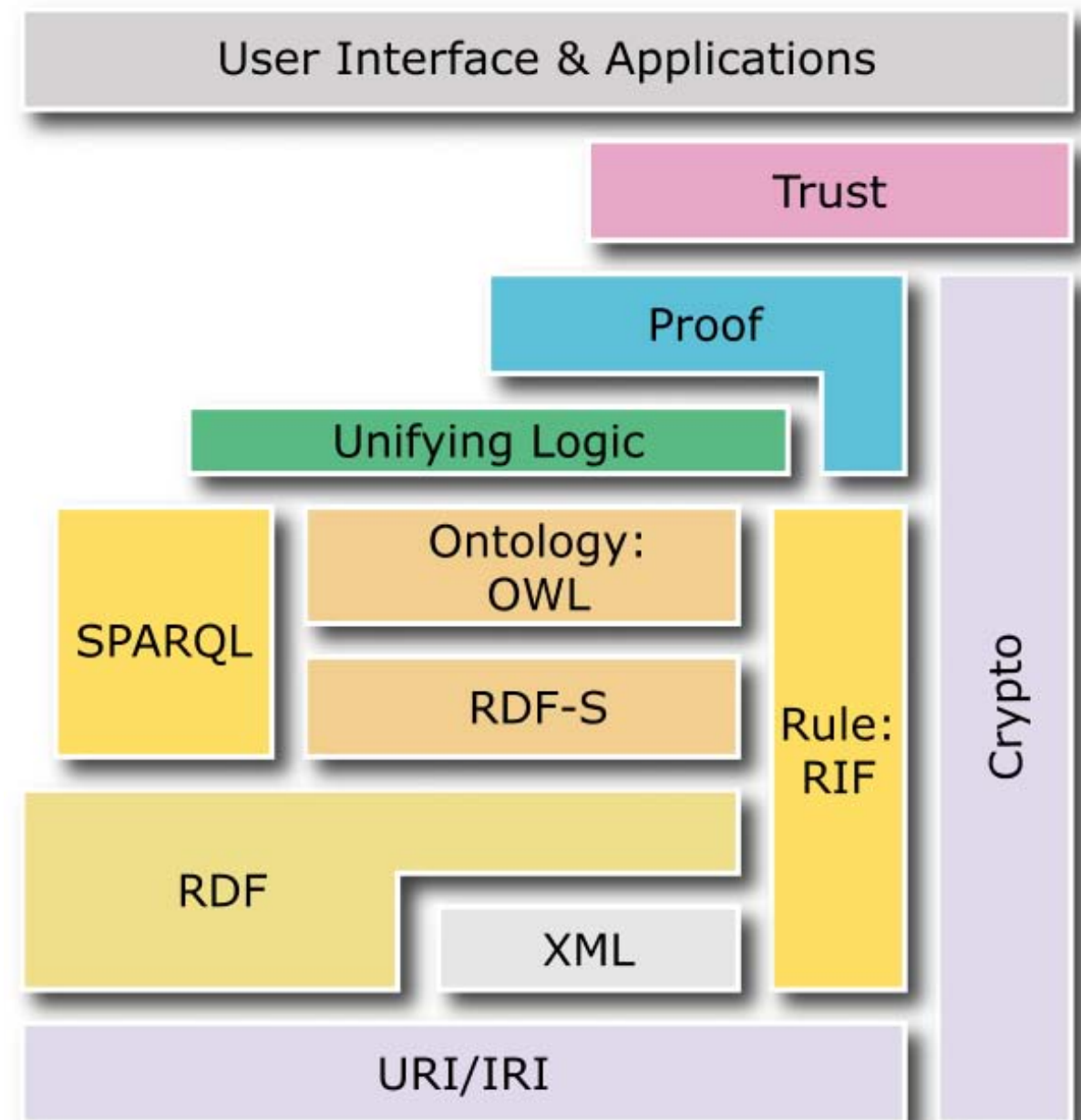


LINKED DATA PRINCIPLES

Semantic technologies



- Semantic Web technologies, standardised by the W3C, are mature
 - RDF recommendation in 1999, update in 2004
 - RDFa (RDF in HTML) note in 2008
 - RDFS recommendation in 2004
 - SPARQL recommendation in 2008
 - OWL recommendation in 2004, update in 2009
- Linked Data is a subset of the Semantic Web stack, including web architecture
 - IRI (IETF RFC 3987, 2005)
 - HTTP (IETF RFC 2616, 1999)



Linked Data principles

AIFB 

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things.

<http://www.w3.org/DesignIssues/LinkedData>

1. Use URIs as names for things

AIFB 

- Use a unique identifier to denote things
- URIs are defined in RFC 2396
- Hegel, Georg Wilhelm Friedrich
 - http://dbpedia.org/resource/Georg_Wilhelm_Friedrich_Hegel
 - <http://viaf.org/viaf/89774942>
 - ...
- Hegel, Georg Wilhelm Friedrich: Gesammelte Werke / Vorlesungen über die Logik
 - [urn:isbn:978-3-7873-1964-0](http://nbn-resolving.org/urn:isbn:978-3-7873-1964-0)



Names for things



*“Now! That should clear up
a few things around here!”*

2. Use HTTP URIs



- Enables “lookup” of URIs
- Via Hypertext Transfer Protocol (HTTP)
- Piggy-backs on hierarchical Domain Name System to guarantee uniqueness of identifiers
- Uses established HTTP infrastructure
- Connects logical level (thing) with physical level (source)
- Important: distinction between name/“thing URI” and location/“source URI” („other resource“/„non-information resource“ vs. „information resource“)

Information resources vs. other resources

AIFB 

Marlene Dietrich, the person

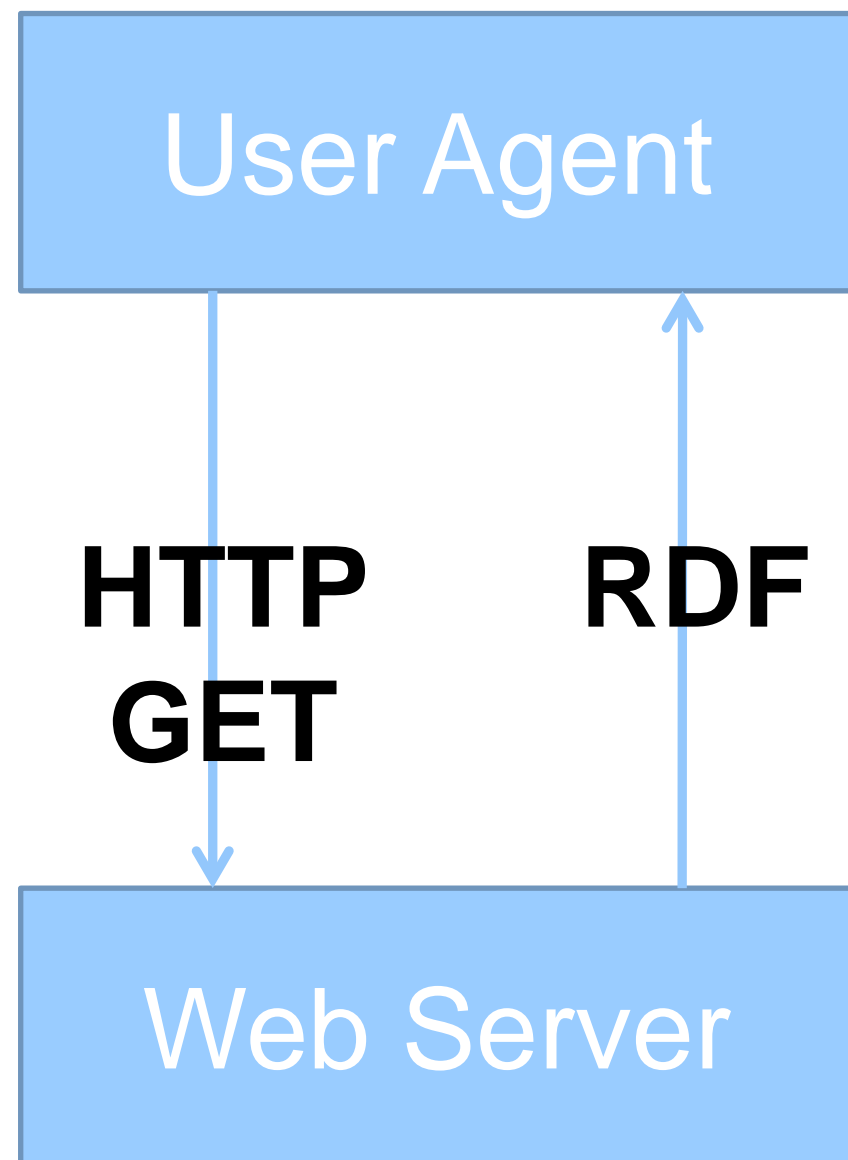


Name?
Creator?
Birth date?
Last change date?
License?
Copyright?
...

File containing data about
Marlene Dietrich



Correspondence between thing-URI and source-URI („hash URIs“)



<http://www.bbc.co.uk/music/artists/191cba6a-b83f-49ca-883c-02b20c7a9dd5#artist>



<http://www.bbc.co.uk/music/artists/191cba6a-b83f-49ca-883c-02b20c7a9dd5.rdf>

Hypertext Transfer Protocol (HTTP)

```
AIFB $ curl -H "Accept: application/rdf+xml" -v  
      http://www.w3.org/People/Berners-Lee/card#i
```

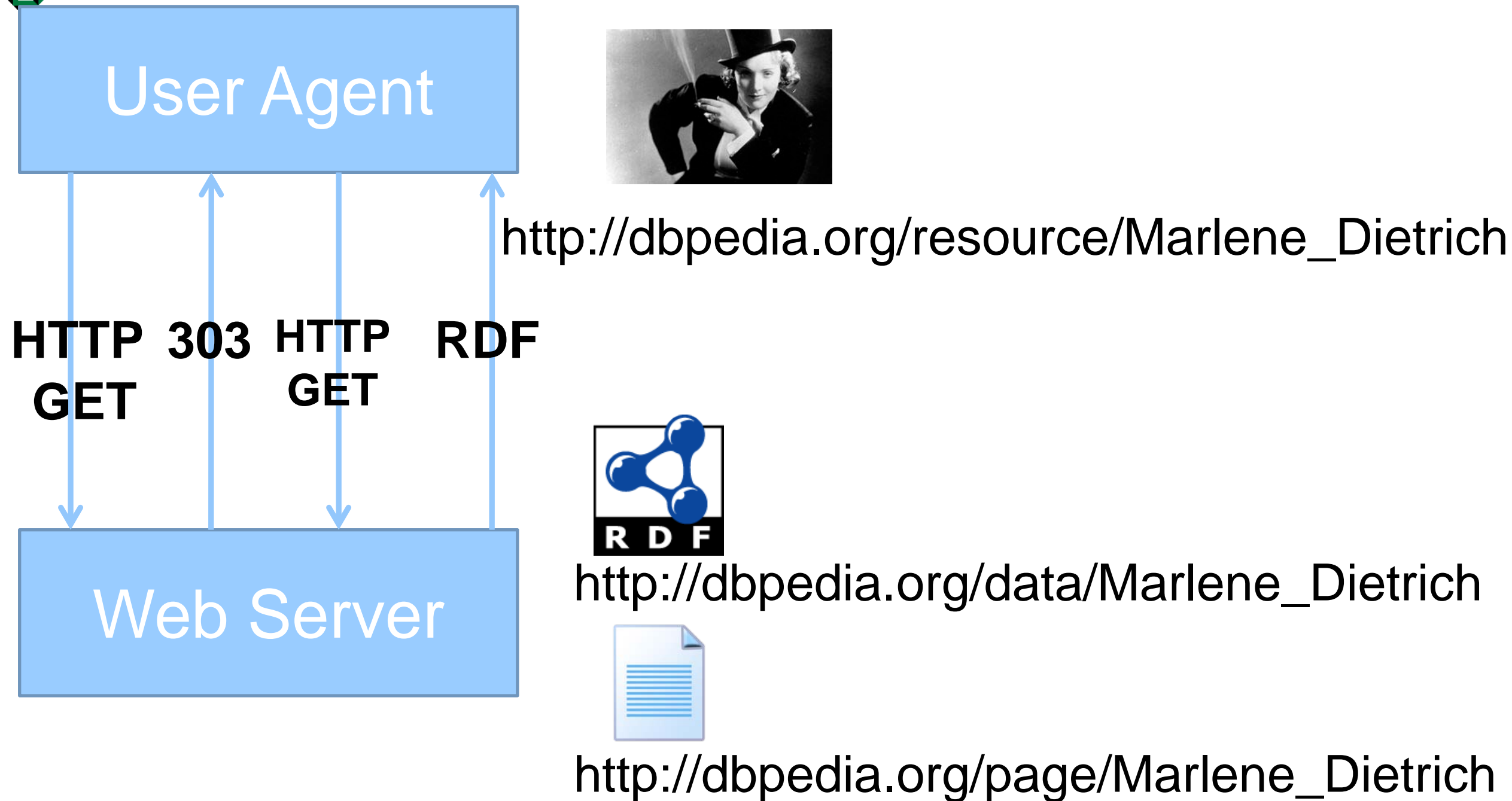
REQUEST

```
> GET /People/Berners-Lee/card HTTP/1.1  
> User-Agent: curl/7.21.0  
> Host: www.w3.org  
> Accept: application/rdf+xml
```

RESPONSE

```
< HTTP/1.1 200 OK  
< Date: Mon, 28 Mar 2011 17:16:30 GMT  
< Server: Apache/2  
< Content-Location: card.rdf
```


Correspondence between thing-URI and source-URI („slash URIs“)

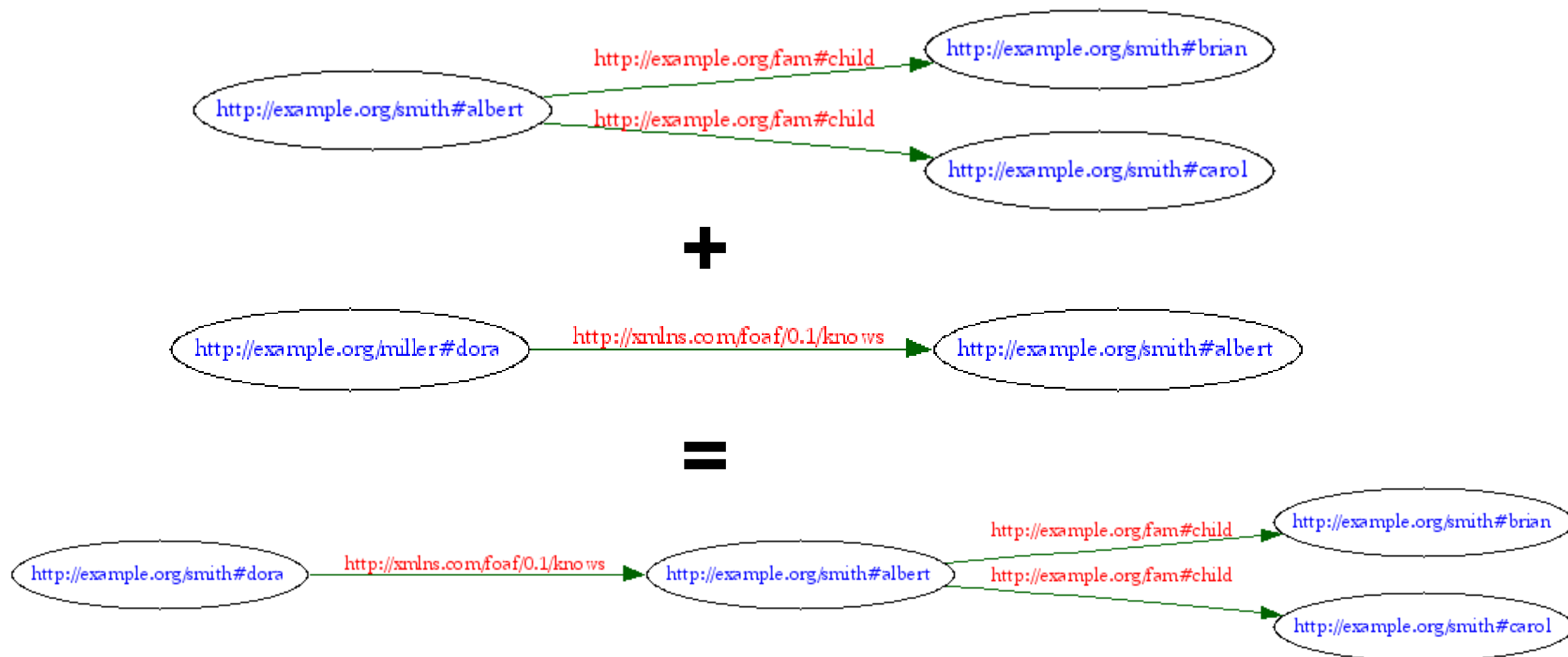
AIFB 

3. Provide useful information

AIFB 

- When somebody looks up a URI, return data using the standards (RDF*, SPARQL)

Merging Data with RDF



4. Link to other URIs

AIFB 

- Enable people (and machines) to jump from server to server
- External links vs. internal links (for any predicate)
- Using external vocabularies enables linking
- Vocabularies might be interlinked, too
- Special owl:sameAs links to denote equivalence of identifiers (useful for data merging)

Equivalences via owl:sameAs



<http://viaf.org/viaf/89774942>

- http://dbpedia.org/resource/Georg_Wilhelm_Friedrich_Hegel
- <http://www.idref.fr/026917467/id>
- <http://libris.kb.se/resource/auth/190350>
- <http://d-nb.info/gnd/118547739>

<http://www.bbc.co.uk/music/artists/191cba6a-b83f-49ca-883c-02b20c7a9dd5#artist>

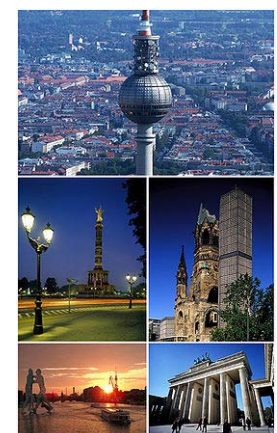
- http://dbpedia.org/resource/Marlene_Dietrich

<http://viaf.org/viaf/97773925>

- http://dbpedia.org/resource/Marlene_Dietrich .
- <http://d-nb.info/gnd/118525565>
- <http://libris.kb.se/resource/auth/238817>
- <http://www.idref.fr/027561844/id>

<http://dbpedia.org/resource/Berlin>

- <http://mpii.de/yago/resource/Berlin>
- <http://data.nytimes.com/N50987186835223032381> - Berlin (Germany)
- <http://www4.wiwiw.fu-berlin.de/flickrwrappr/photos/Berlin>
- <http://data.nytimes.com/16057429728088573361> - Gaspe Peninsula (Quebec) (?)



Benefits of Linked Data



- Explicit, simple data representation
 - Common data representation (Resource Description Framework, RDF) hides underlying technologies and systems
- Distributed system
 - Decentralized distributed ownership and control facilitates adoption and scalability
- Cross-referencing
 - Allows for linking and referencing of existing data, via reuse of URIs
- Loose coupling with common language layer
 - Large scale systems require loose coupling, via HTTP as common access protocol
- Ease of publishing and consumption
 - Simple and easy-to-use systems and technologies to facilitate uptake
- Incremental data integration
 - Start with merged RDF graphs and provide mappings as you go

Challenges

AIFB 

- Ramp-up cost for data conversion
 - May be alleviated by semi-automatic mappings and adequate tool support for manual conversion
- Integrated data may be messy at first
 - But can be refined as need arises
- Distributed creation and loose coordination may result in inconsistencies
 - Can be detected, diagnosed, and fixed with appropriate tools

Data.gov & public sector information

AIFB 

- Many data sets useful for business intelligence



guardian.co.uk

[News](#) | [Sport](#) | [Comment](#) | [Culture](#) | [Business](#) | [Money](#) | [Life & style](#) | [Travel](#) | [Environment](#)
[News](#) > [Datablog](#)

DATABLOG

Facts are sacred

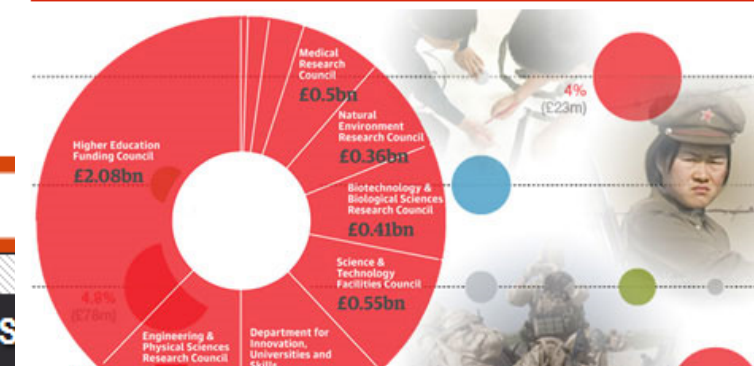
[Previous](#)[Blog home](#)[Next](#)

Data journalism and visualisation: welcome to our new data site

Today we launch a new gateway to our data journalism and visualisations - find out what's new at guardian.co.uk/data

[Tweet](#) 77

[Share](#) 113

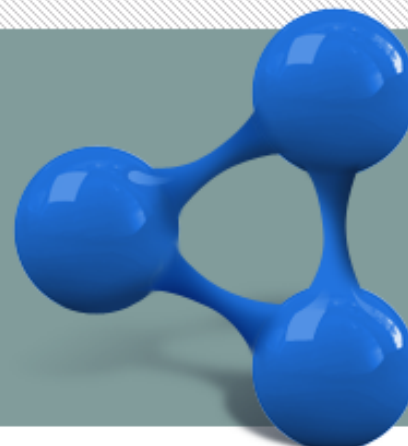
[reddit this](#)
[Comments \(7\)](#)


Posted by
Simon Rogers Thursday
16 December 2010
10.00 GMT
guardian.co.uk

[larger](#) | [smaller](#)

Unlocking innovation

Working with UK Public
Sector information and data



Community
Log in / Sign up

Local Data Panel

the data. A proposal is being
and other formats) to RDF.
nent URI naming with the
riment data (and metadata)

BBC & Media



- Value of content increased by Linked Data

[Home](#)
[News](#)
[Sport](#)
[Weather](#)
[iPlayer](#)
[TV](#)
[Radio](#)
[More...](#)

WILDLIFE FINDER

WATCH THE MOST AMAZING ANIMALS IN THE WORLD

LATEST WILDLIFE

Mosquitoes
 Mosquitoes are a family of insects encompassing over 3,500 recognised species. Most live in warm and humid regions, although they can be found in cooler areas too. While some species feed on nectar, however, the females of some species drink blood to develop eggs.

LION (PANTHERA LEO)

Wildlife Finder > Animals > Lion

SCIENTIFIC CLASSIFICATION

- Kingdom: Animal (animalia)
- ↳ Phylum: Chordate (Chordata)
- ↳ Class: Mammal (Mammalia)
- ↳ Order: Carnivora
- ↳ Family: Felidae
- ↳ Genus: Panthera
- ↳ Species: **Lion (leo)**

OTHER PANTHERA

- Jaguar (onca)
- Leopard (pardus)
- Tiger (tigris)

WAYS OF LIFE

- Acoustic communication
- Carnivorous
- Chemical communication
- Co-operative breeding
- Helpless young
- Kleptoparasitic
- Learning
- Maternal care

COLLECTIONS

Collections offer a new perspective on the vast catalogue of clips showcasing presenter favourites, aspects of wildlife film-making, and gathering content together in illuminating ways.

The wildlife of Life
 In autumn 2009, a major new series brought us life as we've never seen it before. Filmed in HD, the series illustrated the extraordinary lengths to which animals and plants go in order to survive.

David Attenborough's favourite moments
 David Attenborough's favourite moments from the last 30 years represent not only memorable personal experiences... but also the...

EDITOR'S CHOICE

This week's highlights feature the wild stars of Life and Natural World.

 Sunflower star 2 videos	 Chimpanzee 6 videos
 Giant cuttlefish 2 videos	 Hippopotamus 6 videos
 Amphibian 2 videos	 Bird 2 videos

WHAT'S NEW?

30 November 2009
Homepage redesign: [subscribe](#) to the RSS feed for new content alerts

Browse the image carousel to see the latest wildlife pages

Check out the images of fish that look like...

Semantic technologies at BBC



- Various micro-sites built and maintained manually
- No integration across sites in terms of content and metadata
- Use cases
 - Find and explore content on specific (and related) topics
 - Maintain and re-organize sites
 - Leverage external resources
- Ontology: One page per thing, reusing DBpedia and MusicBrainz IDs, different labels...

The screenshot shows the BBC Music BETA website interface. At the top, there's a navigation bar with 'BBC', 'Text only', and 'Help'. Below it, the 'Music BETA' header is visible. The main content area features a large image of a performer and a section titled 'Most Played Artists On The BBC' for the period '30 NOVEMBER - 6 DECEMBER 2009'. A horizontal list of artists is displayed with their play counts: 1. Dizzee Rascal, 2. Johnny Cash, 3. Mumford & Sons, 4. Jason Derulo, 5. Arctic Monkeys, 6. Donae'o, 7. Jay. Below this list, there's a note about incomplete information and a link to 'Find out more about this artist play count information.' Further down, there are sections for 'Latest News Stories' and 'Music Radio Programmes'.

Overlaid on the right side of the screenshot is a network diagram. It features several interconnected nodes: 'Programmes' (green), 'Music' (orange), 'Events' (pink), 'Topics' (yellow), and 'Users' (purple). External nodes include 'Programmes Ontology', 'PIPS', 'Music Ontology', 'MusicBrainz', 'CIS', 'SKOS', 'Event Ontology', 'OpenID?', and 'FOAF'. The central node 'BBC' is connected to all the main nodes.

„Design for a world where Google is your homepage, Wikipedia is your CMS, and humans, software developers and machines are your users“

<http://www.slideshare.net/reduxd/beyond-the-polar-bear>

Semantic technologies at BBC(2)



Text only | Help

BBC Home News Sport Weather iPlayer TV Radio More... Search

SPORT **WORLD CUP 2010**

SPORT FOOTBALL WORLD CUP 2010 GROUPS & TEAMS FIXTURES & RESULTS VIDEO BBC COVERAGE

Latest matches

England

England 1-1 United States Saturday, 12 June Match report

England 0-0 Algeria Friday, 18 June Match report

Slovenia 0-1 England Wednesday, 23 June Match report

Germany 4-1 England Sunday, 27 June Match report

Latest stories

Gerrard commits future to England **NEW**

• England sponsorship likely to end

• Capello to remain England manager

• Mueller blames England imbalance

• Capello receives Gartside backing

Pressure got to Rooney - Ferguson

• FA unfit for purpose says Caborn

• England's fear of crossing borders

• England duo bypass London event

• Barwick baffled by dismal England

German lessons

Jurgen Klinsmann on how to revolutionise England

- A German view on English football
- Redknapp backs England to shine
- BBC pundits on England
- Roy Hodgson Q&A
- World Cup goals analysis

Around the web

- BBC Search+ counts
- England Fifa

